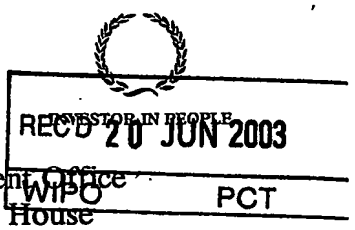
PCT/GB 2003/001340



The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

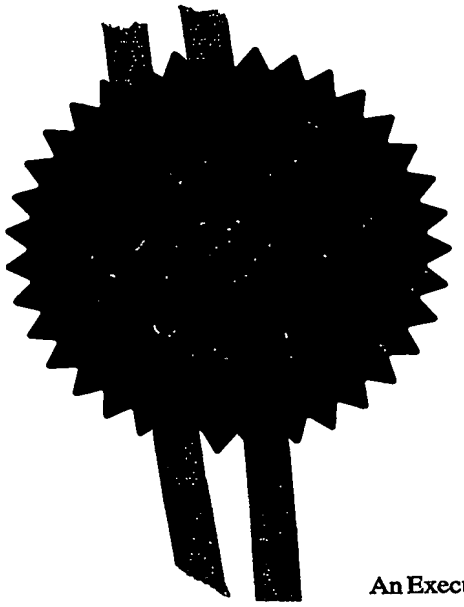
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

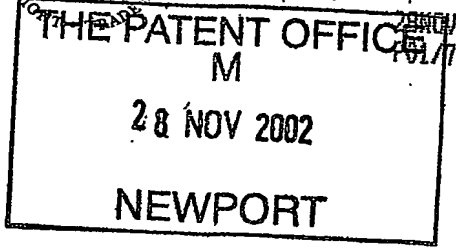


Signed Stephen Hordley

Dated 9 April 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)



1/77

28NOV02 E766928-6 D00239
P01/7700 0.00-0227733.3

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

1. Your reference PC/JM/P12562GB

2. Patent application number *(The Patent Office will fill in this part)* 0227733.3

28 NOV 2002

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

UNIVERSITY COURT OF THE UNIVERSITY OF EDINBURGH
Old College
South Bridge
Edinburgh EH8 9YL, UNITED KINGDOM

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation 7986780001

4. Title of the invention

Schizophrenia Associated Gene (V)

5. Name of your agent *(if you have one)*

CRUIKSHANK & FAIRWEATHER

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

19 Royal Exchange Square
Glasgow G1 3AE
UNITED KINGDOM

Patents ADP number *(if you know it)* 547002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	22 ✓
Claim(s)	–
Abstract	–
Drawing(s)	5 + 5 ✓

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Cruikshank & Fairweather Date 27/11/02.

CRUIKSHANK & FAIRWEATHER 27 NOVEMBER 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

DR. P. CHAPMAN - 0141-221-5767

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

SCHIZOPHRENIA ASSOCIATED GENE (V)

The present invention relates to the identification of a gene which has been disrupted in a patient diagnosed as suffering from schizoaffective disorder, as well as proteins encoded by the gene and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis, such as schizophrenia.

Schizophrenia and Bipolar Affective Disorder are common and debilitating psychiatric disorders. Despite a wealth of information on the epidemiology, neuroanatomy and pharmacology of the illness, it is uncertain what molecular pathways are involved and how impairments in these affect brain development and neuronal function. Despite an estimated heritability of 60-80%, very little is known about the number or identity of genes involved in these psychoses. Although there has been recent progress in linkage and association studies, especially from genome-wide scans, these studies have yet to progress from the identification of susceptibility loci or candidate genes to the full characterisation of disease-causing genes (Berrettini, 2000).

The cloning of breakpoints in patients with chromosome abnormalities (translocations, inversions etc.) has proved instrumental in the identification of many disease genes (e.g. Duchenne Muscular Dystrophy, Retinoblastoma, Wilm's Tumour, Familial Polyposis Coli, Fragile-X Syndrome, Polycystic Kidney Disease, many leukaemias and, very recently, a candidate speech and language disorder gene (Lai et al, 2001)). Such studies assume that the chromosomal breakpoints give rise to the clinical symptoms by either directly disrupting

gene sequences or perturbing gene expression. In the same way that gene-trap mutagenesis can be used to identify disrupted mouse genes (Brennan & Skarnes, 1999), the physical "flag" created by a cytogenetic breakpoint provides a geographical pointer for the disease locus.

It is amongst the objects of the present invention to provide a gene and/or protein postulated to be involved with the development and/or symptoms associated with schizophrenia and/or affective psychosis.

As will be seen, the present invention is based on the molecular characterisation of a chromosomal rearrangement denoted t(3;8)(p13;p22) in a subject diagnosed as suffering from a schizoaffective disorder (see Fig.1). A high-throughput Fluorescence *in situ* Hybridisation (FISH)-based approach was adopted to map the chromosomal breakpoints in these patients. Consultation of the sequence data at the breakpoint loci not only allowed efficient FISH probe selections to be made by the targeting of coding regions, but also proof of gene disruption was inferred entirely by relating the exact position of probes to the genomic structure of a candidate gene.

One breakpoint (located on chromosome 8p22) in this subject lies near to a gene, *N33*, involved in the N-Linked Glycosylation pathway. Without wishing to be bound by theory it is hypothesised that the breakpoint in the subject perturbs *N33* expression indirectly through position effect silencing or separation of regulatory elements from the gene promoter (both effects have been shown to occur even when the breakpoints are up to 1Mb from the target gene in some instances (Kleinjan et al 1999)). The *N33* gene is located within a chromosomal region repeatedly found positive in schizophrenia linkage studies. This gene is the subject of filed patent GB0207902.8.

The other breakpoint in this patient (3p13) has now been fully characterised and demonstrated to disrupt a gene, *SEMCAP3* (also known as KIAA1095). The present invention is therefore based on a proposed role of this gene (normal and mutated forms) in the aetiology of schizophrenia and/or affective psychosis.

Thus, in a first aspect the present invention provides use of a polynucleotide fragment comprising the *SEMCAP3* gene, or a fragment, derivative or homologue thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

In another aspect, the present invention provides use of a polypeptide fragment encoded by the *SEMCAP3* gene, or fragments, derivatives or homologues thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

Schizophrenia and/or affective psychosis as used herein relates to schizophrenia, as well as other affective psychoses such as those listed in "The ICD-10 Classification of Mental and Behavioural Disorders" World Health Organization, Geneva 1992. Categories F20 to F29 inclusive includes Schizophrenia, schizotypal and delusional disorders. Categories F30 to F39 inclusive are Mood (affective) disorders that include bipolar affective disorder and depressive disorder. Mental Retardation is coded F70 to F79 inclusive. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). American Psychiatric Association, Washington DC. 1994. Include all conditions coded 295.xx (Schizophrenia and Other Psychotic Disorders) and 296.xx (Major Depressive Disorders and Bipolar Disorders). Mental Retardation is coded 315, 317, 318 and 319.

SEMCAP3 has been previously cloned and sequenced in mouse as two alternative forms (Semcap3*A* and *3B*) and the sequences are present in the public database (nucleic acid sequences; AF127084/AF127085, respectively; protein sequences AAF22131/AAF22132,

respectively) as directly submitted by Wang & Strittmatter, 1999. The human form of the gene is defined by sequence KIAA1095 (nucleic acid sequence, AB029018 or XM_041363, and a smaller form, BC014432; protein sequence, XP_041363). The genomic sequences corresponding to this gene are also present in the public database (eg. for BAC RP11-252o10, AC024102). Nevertheless, the prior art does not suggest any link between *SEMCAP3* and schizophrenia and/or affective psychosis.

Thus, references herein to the *SEMCAP3* gene are understood to relate to the sequences in the public databases and identified in Fig.3 and references to the SEMCAP3 protein sequence is understood to relate to the sequences in the public databases and identified in Fig.4.

In certain jurisdictions claims to methods of treatment are permissible and so the skilled reader will appreciate that the *SEMCAP3* gene, or fragments, derivatives or homologues thereof; or SEMCAP3 protein, or functionally active fragments, derivatives, or homologues thereof, may be administered to an individual as a method of treating an individual with schizophrenia and/or affective psychosis.

"Polynucleotide fragment" as used herein refers to a chain of nucleotides such as deoxyribose nucleic acid (DNA) and transcription products thereof, such as RNA. Naturally, the skilled addressee will appreciate the whole naturally occurring human genome is not included in the definition of polynucleotide fragment.

The polynucleotide fragment can be isolated in the sense that it is substantially free of biological material with which the whole genome is normally associated *in vivo*. The isolated polynucleotide fragment may be cloned to provide a recombinant molecule comprising the polynucleotide fragment. Thus, "polynucleotide fragment includes double and single stranded DNA, RNA and polynucleotide sequences derived therefrom, for example, subsequences of said fragment

and which are of any desirable length. Where a nucleic acid is single stranded then both a given strand and a sequence or reverse complementary thereto is within the scope of the present invention.

In general, the term "expression product" or "gene product" refers to both transcription and translation products of said polynucleotide fragments. When the expression or gene product is a "polypeptide" (i.e. a chain or sequence of amino acids displaying a biological activity substantially similar (eg. 98%, 95%, 90%, 80%, 75% activity) to the biological activity of the protein), it does not refer to a specific length of the product as such. Thus, the skilled addressee will appreciate that "polypeptide" encompasses *inter alia* peptides, polypeptides and proteins. The polypeptide if required, can be modified *in vivo* and *in vitro*, for example by glycosylation, amidation, carboxylation, phosphorylation and/or post-translational cleavage.

The present invention further provides a recombinant or synthetic polypeptide for the manufacture of reagents for use as therapeutic agents in the treatment of schizophrenia and/or affective psychosis. In particular, the invention provides pharmaceutical compositions comprising the recombinant or synthetic polypeptide together with a pharmaceutically acceptable carrier therefor.

The present invention further provides an isolated polynucleotide fragment capable of specifically hybridising to a related polynucleotide sequence from another species. In this manner, the present invention provides probes and/or primers for use in *ex vivo* and/or *in situ* detection and expression studies. Typical detection studies include polymerase chain reaction (PCR) studies, hybridisation studies, or sequencing studies. In principle any specific polynucleotide sequence fragment from the identified sequences may be used in detection and/or expression studies. The skilled addressee understands that a

specific fragment is a fragment of the sequence which is of sufficient length, generally greater than 10, 12, 14, 16 or 20 nucleotides in length, to bind specifically to the sequence, under conditions of high stringency, as defined herein, and not bind to unrelated sequences, that is sequences from elsewhere in the genome of the organism other than an allelic form of the sequence or non-homologous sequences from other organisms.

"Capable of specifically hybridising" is taken to mean that said polynucleotide fragment preferably hybridises to a related or similar polynucleotide sequence in preference to unrelated or dissimilar polynucleotide sequences.

The invention includes polynucleotide sequence(s) which are capable of specifically hybridising to a polynucleotide fragment as described herein or to a part thereof without necessarily being completely complementary or reverse complementary to said related polynucleotide sequence or fragment thereof. For example, there may be at least 50%, or at least 75%, at least 90%, or at least 95% complementarity. Of course, in some cases the sequences may be exactly reverse complementary (100% reverse complementary) or nearly so (e.g. there may be less than 10, typically less than 5 mismatches). Thus, the present invention also provides anti-sense or complementary nucleotide sequence(s) which is/are capable of specifically hybridising to the disclosed polynucleotide sequence. If a specific polynucleotide is to be used as a primer in PCR and/or sequencing studies, the polynucleotide must be capable of hybridising to related nucleic acid and capable of initiating chain extension from 3' end of the polynucleotide, but not able to correctly initiate chain extension from unrelated sequences.

If a polynucleotide sequence of the present invention is to be used in hybridisation studies to obtain or identify a related sequence from another organism the polynucleotide sequence should preferably

remain hybridised to a sample polynucleotide under stringent conditions. If desired, either the test or sample polynucleotide may be immobilised. Generally the test polynucleotide sequence is at least 10, 14, 20 or at least 50 bases in length. It may be labelled by suitable techniques known in the art. Preferably the test polynucleotide sequence is at least 200 bases in length and may even be several kilobases in length. Thus, either a denatured sample or test sequence can be first bound to a support. Hybridization can for example be effected at a temperature of between 50 and 70°C in double strength SSC (2xNaCl 17.5g/l and sodium citrate (SC) at 8.8g/l) buffered saline containing 0.1% sodium dodecyl sulphate (SDS). This can be followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, such reduced concentration buffers are typically single strength SSC containing 0.1%SDS, half strength SSC containing 0.1%SDS and one tenth strength SSC containing 0.1%SDS. Sequences having the greatest degree of similarity are those the hybridisation of which is least affected by washing in buffers of reduced concentration. It is most preferred that the sample and inventive sequences are so similar that the hybridisation between them is substantially unaffected by washing or incubation in standard sodium citrate (0.1 x SSC) buffer containing 0.1%SDS.

Oligonucleotides may be designed to specifically hybridise to *SEMCAP3* nucleic acids. They may be synthesised, by known techniques and used as primers in PCR or sequencing reactions or as probes in hybridisations designed to detect the presence of a mutated or normal *SEMCAP3* gene in a sample. The oligonucleotides may be labelled by suitable labels known in the art, such as, radioactive labels, chemiluminescent labels or fluorescent labels and the like. The term "oligonucleotide" is not meant to

indicate any particular length of sequence and encompasses nucleotides of preferably at least 10b (e.g. 10b to 1kb) in length, more preferably 12b-500b in length and most preferably 15b to 100b.

The oligonucleotides may be designed with respect to any of the sequences described herein and may be manufactured according to known techniques. They may have substantial sequence identity (e.g. at least 50%, at least 75%, at least 90% or at least 95% sequence identity) with one of the strands shown herein or an RNA equivalent, or with a part of such a strand. Preferably such a part is at least 10, at least 30, at least 50 or at least 200 bases long. It may be an open reading frame (ORF) or a part thereof.

Oligonucleotides which are generally greater than 30 bases in length should preferably remain hybridised to a sample polynucleotide under one or more of the stringent conditions mentioned above. Oligonucleotides which are generally less than 30 bases in length should also preferably remain hybridised to a sample polynucleotide but under different conditions of high stringency. Typically the melting temperature of an oligonucleotide less than 30 bases may be calculated according to the formula of; 2°C for every A or T, plus 4°C for every G or C, minus 5°C. Hybridization may take place at or around the calculated melting temperature for any particular oligonucleotide, in 6 x SSC and 1% SDS. Non specifically hybridised oligonucleotides may then be removed by stringent washing, for example in 3 x SSC and 0.1% SDS at the same temperature. Only substantially similar matched sequences remain hybridised i.e. said oligonucleotide and corresponding test nucleic acid.

When oligonucleotides of generally less than 30 bases in length are used in sequencing and/or PCR studies, the melting temperature may be calculated in the same manner as described above. The oligonucleotide may then be allowed to anneal or hybridise at a temperature around the oligonucleotides

calculated melting temperature. In the case of PCR studies the annealing temperature should be around the lower of the calculated melting temperatures for the two priming oligonucleotides. It is to be appreciated that the conditions and melting temperature calculations are provided by way of example only and are not intended to be limiting. It is possible through the experience of the experimenter to vary the conditions of hybridisation and thus anneal/hybridise oligonucleotides at temperatures above their calculated melting temperature. Indeed this can be desirable in preventing so-called non-specific hybridisation from occurring.

It is possible when conducting PCR studies to predict an expected size or sizes of PCR product(s) obtainable using an appropriate combination of two or more oligonucleotides, based on where they would hybridise to the sequences described herein. If, on conducting such a PCR on a sample of DNA, a fragment of the predicted size is obtained, then this is predictive that the DNA encodes a homologous sequence from a test organism.

Proteins for all the applications described herein can be produced by cloning the gene for example into plasmid vectors that allow high expression in a system of choice e.g. insect cell culture, yeast, animal cells, bacteria such as *Escherichia coli*. To enable effective purification of the protein, a vector may be used that incorporates an epitope tag (or other "sticky" extension such as His6) onto the protein on synthesis. A number of such vectors and purification systems are commercially available.

The polynucleotide fragment can be molecularly cloned into a prokaryotic or eukaryotic expression vector using standard techniques and administered to a host. The expression vector is taken up by cells and the polynucleotide fragment of interest expressed, producing protein.

It will be understood that for the particular polypeptides embraced herein, natural variations such as may occur due to polymorphisms, can exist between individuals or between members of the family. These variations may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. All such derivatives showing the recognised activity are included within the scope of the invention. For example, for the purpose of the present invention conservative replacements may be made between amino acids within the following groups:

(I) Alanine, serine, threonine;

(II) Glutamic acid and aspartic acid;

(III) Arginine and leucine;

(IV) Asparagine and glutamine;

(V) Isoleucine, leucine and valine;

(VI) Phenylalanine, tyrosine and tryptophan

Moreover, recombinant DNA technology may be used to prepare nucleic acid sequences encoding the various derivatives outlined above.

As is well known in the art, the degeneracy of the genetic code permits substitution of bases in a codon resulting in a different codon which is still capable of coding for the same amino acid, e.g. the codon for amino acid glutamic acid is both GAT and GAA. Consequently, it is clear that for the expression of polypeptides from nucleotide sequences described herein or fragments thereof, use can be made of a derivative nucleic acid sequence with such an alternative codon composition different from the nucleic acid sequences shown in the Figures.

The polynucleotide fragments of the present invention are preferably linked to regulatory control sequences. Such control sequences may comprise promoters, operators, inducers, enhancers, silencers, ribosome binding sites, terminators etc. Suitable

control sequences for a given host may be selected by those of ordinary skill in the art.

A polynucleotide fragment according to the present invention can be ligated to various expression controlling sequences, resulting in a so called recombinant nucleic acid molecule. Thus, the present invention also includes an expression vector containing an expressible nucleic acid molecule. The recombinant nucleic acid molecule can then be used for the transformation of a suitable host.

Specific vectors which can be used to clone nucleic acid sequences according to the invention are known in the art (e.g. Rodriguez, R.L. and Denhadt, D.T., Edit., Vectors: a survey of molecular cloning vectors and their uses, Butterworths, 1988, or Jones *et al.*,Vectors: Cloning Applications: Essential Techniques (Essential techniques series), John Wiley & Son. 1998).

The methods to be used for the construction of a recombinant nucleic acid molecule according to the invention are known to those of ordinary skill in the art and are inter alia set forth in Sambrook, et al. (Molecular Cloning: a laboratory manual Cold Spring Harbour Laboratory, 1989).

The present invention also relates to a transformed cell containing the polynucleotide fragment in an expressible form. "Transformation", as used herein, refers to the introduction of a heterologous polynucleotide fragment into a host cell. The method used may be any known in the art, for example, direct uptake, transfection transduction or electroporation (Current Protocols in Molecular Biology, 1995. John Wiley and Sons Inc.). The heterologous polynucleotide fragment may be maintained through autonomous replication or alternatively, may be integrated into the host genome. The recombinant nucleic acid molecules preferably are provided with appropriate control sequences compatible with the designated host which can regulate the expression of the inserted polynucleotide fragment, e.g. tetracycline responsive

promoter, thymidine kinase promoter, SV-40 promoter and the like.

Suitable hosts for the expression of recombinant nucleic acid molecules may be prokaryotic or eukaryotic in origin. Hosts suitable for the expression of recombinant nucleic acid molecules may be selected from bacteria, yeast, insect cells and mammalian cells.

In another aspect the present invention also relates to a method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affective psychosis in an individual, wherein the method comprises determining if the *SEMCAP3* gene in the individual has been disrupted by a mutation or chromosomal rearrangement.

The methods which may be employed to elucidate such a mutation or chromosomal rearrangement are well known to those of skill in the art and could be detected for example using PCR or in hybridisation studies using suitable probes which could be designed to span an identified mutation site or chromosomal breakpoint in close proximity to the *SEMCAP3* gene, such as the breakpoint identified by the present inventors and described herein.

Once a particular polymorphism or mutation has been identified it may be possible to determine a particular course of treatment. For example it is known that some forms of treatment work for some patients, but not all. This may in fact be due to mutations in the *SEMCAP3* gene or surrounding sequences, and it may therefore be possible to determine a treatment strategy using current therapies, based on a patient's genotype.

It will be appreciated that mutations in the gene sequence or controlling elements of a gene, eg. a promoter and/or enhancer can have subtle effects such as affecting mRNA splicing/stability/activity and/or control of gene expression levels, which can also be determined. Also the relative levels of RNA can be determined using for example hybridisation or

quantitative PCR as a means to determine if the *SEMCAP3* gene has been disrupted.

Moreover the presence and/or levels of the *SEMCAP3* gene product itself can be assayed by immunological techniques such as radioimmunoassay, Western blotting and ELISA using specific antibodies raised against the gene products. The present invention also therefore relates to antibodies specific for *SEMCAP3* gene product and uses thereof in diagnosis and/or therapy.

Thus, in a further aspect of the present invention provides antibodies specific to the polypeptides of the present invention or epitopes thereof. Production and purification of antibodies specific to an antigen is a matter of ordinary skill, and the methods to be used are clear to those skilled in the art. The term antibodies can include, but is not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanised or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope binding fragments of any of the above. Such antibodies may be used in modulating the expression or activity of the particular polypeptide, and/or in detecting said polypeptide *in vivo* or *in vitro* and thus used, for example in the treatment and/or diagnosis of schizophrenia and/or affective psychosis.

Using the sequences disclosed herein, it is possible to identify related sequences in other animals, such as mammals, with the intention of providing an animal model for psychiatric disorders associated with the improper functioning of the nucleotide sequences and proteins of the present invention. Once identified, the homologous sequences can be manipulated in several ways common to the skilled person in order to alter the functionality of the nucleotide sequences and proteins homologous to those of the present invention. For example, "knock-out" animals may be created, that is, the expression of

the genes comprising the nucleotide sequences homologous to those of the present invention may be reduced or substantially eliminated in order to determine the effects of reducing or substantially eliminating the expression of such genes. Alternatively, animals may be created where the expression of the nucleotide sequences and proteins homologous to those of the present invention are upregulated, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be increased in order to determine the effects of increasing the expression of these genes. In addition to these manipulations, substitutions, deletions and additions may be made to the nucleotide sequences encoding the proteins homologous to those of the present invention in order to effect changes in the activity of the proteins to help elucidate the function of domains, amino acids, etc. in the proteins. Furthermore, the sequences of the present invention may also be used to transform animals to the manner described above. The manipulations described above may also be used to create an animal model of schizophrenia and/or affective psychosis associated with the improper functioning of the nucleotide sequences and/or proteins of the present invention in order to evaluate potential agents which may be effective for combatting psychotic disorders, such as schizophrenia and/or affective psychosis.

Thus, the present invention also provides for screens for identifying agents suitable for preventing and/or treating schizophrenia and/or affective psychosis associated with disruption or alteration in the expression of the *SEMCAP3* gene and/or its gene products. Such screens may easily be adapted to be used for the high throughput screening of libraries of compounds such as synthetic, natural or combinatorial compound libraries.

Thus, the *SEMCAP3* gene product according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with nucleotide fragments according to the invention or an expression vector comprising a polynucleotide fragment according to the invention, said cells expressing the *SEMCAP3* gene product according to the invention.

Alternatively also the *SEMCAP3* gene product according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the *SEMCAP3* gene product.

Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or inhibition of a functional response is measured.

Thus, the present invention provides for a method for identifying ligands for *SEMCAP3* gene product, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

Compounds which activate or inhibit the function of *SEMCAP3* gene product may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present invention.

The present invention will now be further described by way of Example and with reference to the Figures which show:

Figure 1 shows an ideogram diagram of the chromosomal rearrangement (a reciprocal translocation) in this patient. The two breakpoints are marked at the approximate chromosomal locations at which they are located. In addition, and not to scale, the two candidate disease-causing genes, *N33* and *SEMCAP3*, are placed in the correct orientation and with respect to the breakpoints.

Figure 2 shows a representation of the genomic structure of the *SEMCAP3* gene: its spliced exons spread over a genomic extent of approximately 250kb.
Above the gene, the coding contribution of each exon to the SEMCAP3 protein is indicated by bars and finely dashed lines. The domain structure of SEMCAP3 protein is shown at the top of the figure. `RING' refers to a RING-finger domain, `ZF-T.' to a TRAF-type zinc finger (also referred to as a *sina* domain) and `PDZ' to PDZ domain present in PSD-95, Dlg, and ZO-1/2. The BAC clones used to identify the breakpoint location are included at the bottom of the figure together with the inferred direction (arrows) of the breakpoint from the FISH results using these clones. The heavy dashed line shows the position of the breakpoint with respect to the gene exons and the domain structure of the protein.

Figure 3 Nucleic acid sequence of Human *SEMCAP3* (genomic DNA sequence including CpG island/putative promoter upstream of 5' UTR/cDNA sequence is also included for clarity). The following features are marked for clarity:

a) ATG start site located at position 709 (underlined)
b) GG bases (underlined) at the junction between exons 3 and 4 (i.e. between which the breakpoint is located)

c) UAA stop codon located at position 3907 (underlined).

Figure 4 Amino acid sequence of Human SEMCAP3 with underlined regions of interest.

a) Residues 18-55 Ring finger domain
b) Residues 101-158 SINA/ZF-TRAF domain
c) Residues 246-339 PDZ domain #1
d) Residues 418-504 PDZ domain #2

Materials and methods

Lymphocyte extraction and metaphase chromosome preparation

Lymphocytes were extracted from 7mls of patient blood (for storage and generation of EBV-transformed cell lines) using density gradient separation (Histopaque-1077, Sigma). In order to generate metaphase-arrested chromosomes for cytogenetic analysis, 0.8mls of patient blood were cultured for 71hrs in medium containing phytohaemagglutinin (Peripheral Blood Medium, Sigma). The short-term cultures were treated with colcemid for one hour followed by a conventional fixing procedure. Fixed chromosomes were dropped onto microscope slides and stored for 1 week prior to use in FISH experiments.

Fluorescence in situ hybridisation (FISH) protocol

Probe template DNA (e.g BAC clone DNA) was labelled by nick translation and hybridised to patient metaphase spreads using standard FISH methods. Slides were counterstained with DAPI in Vectashield anti-fade solution (Vector laboratories). A Zeiss Axioskop fluorescence microscope with a chroma number 83000 multi-spectral filter set was used to observe the chromosomal hybridisations. Images were captured using Digital Scientific SmartCapture imaging software. FISH signals observed on derived chromosomes dictated the selection of further clones required to "walk" towards the breakpoint. Breakpoint-spanning FISH probes have

signals on a normal chromosome and on both derived chromosomes.

Resolution of breakpoint position

BAC clones corresponding to proposed breakpoint regions were arranged into contigs by consulting the Washington University FPC (http://www.genome.wustl.edu/gsc/human/Mapping/index.shtml), UCSC GoldenPath Draft Human Genome Browser (http://genome.cse.ucsc.edu/index.html?org=Human) and Ensembl (http://www.ensembl.org/) databases. BAC clones were supplied by BACPAC Resources, Oakland, California, USA (http://www.chori.org/bacpac/). Clone selection was biased to gene-containing BACs.

Example 1: Molecular characterisation of chromosomal disruption and identification of disrupted gene

FISH experiments on chromosome 3p13 had narrowed the location of the breakpoint to a region including the large gene *SEMCAP3* (approximately 250kb genomic extent). Two BAC clones were selected from the tiling diagram of BAC clones placed on the human genome map backbone (June 2002 release of the `BAC End Pairs' track on the UCSC Genome Browser; http://genome.cse.ucsc.edu/index.html?org=Human). These were RPCI-11 606p16 and RPCI-11 94j25. By FISH, these BAC clones flanked the breakpoint (the former translocated to the derived chromosome 8 and the latter remained on the derived chromosome 3). The position of these two BAC clones indicated that the breakpoint lay within the large (200kb) intron between exons 3 and 4 of the *SEMCAP3* gene (see Fig.2). Thus, the inventors inferred from these results that the *SEMCAP3* gene was directly disrupted by the 3p13 translocation event and, as such, is a candidate gene for the psychiatric disorder exhibited by the patient.

Semcap3 (semaphorin cytoplasmic domain-associated protein) was originally identified in mouse as a gene encoding a protein that interacts with M-semF/Sema4c. Two forms, 3A and 3B, were submitted to the public nucleic acid sequence database (Wang & Strittmatter, 1999) but have yet to be published. It appears that 3b may be an artifactual sequence as it displays deletions in the sequence. Sema3a is identical in structure to the predicted human gene, KIAA1095 and the inventors refer to this sequence as human *SEMCAP3*. The yeast two-hybrid screen that isolated Sema3a/b also identified Sema1 and Sema2 as genes encoding proteins which interact with the cytoplasmic tail of the SEMA4C protein (Wang et al., 1999).

The purpose of these screening experiments was to elucidate cytoplasmic interactors with the transmembrane receptor, SEMA4C. This protein belongs to a large group of signalling proteins described as `semaphorins'. In the brain, these proteins are thought to play important roles in brain development through their action on axonal guidance and growth cone stability. Inagaki et al., (1995) showed that Sema4C is expressed in the developing mouse brain. One proposed explanation for the origin of psychiatric disorders (including the disorder exhibited by the patient described here) is the incorrect development of the brain, particularly the connections, projections and neural networks between brain subregions. With this in mind, semaphorins, and the proteins that interact with them (such as the SEMCAPs), become attractive candidate genes for the psychiatric disorders.

It is suspected that the PDZ domains (see Fig.2) of the SEMCAP3 protein will be involved in protein-protein interactions (such as SEMA4C interaction) as they are in other proteins. The RING-finger domain of SEMCAP3 identifies it as belonging to a class of proteins known as ubiquitin ligases. Ubiquitin ligases specifically target proteins for ubiquitination and subsequent destruction in the proteasome pathway. Thus,

SEMCAP3 may act to regulate the activity of other proteins (for instance, components of the semaphorin pathway) by targeting them for destruction. The ZF-TRAF/SINA domain is most likely an extension of the RING-finger domain.

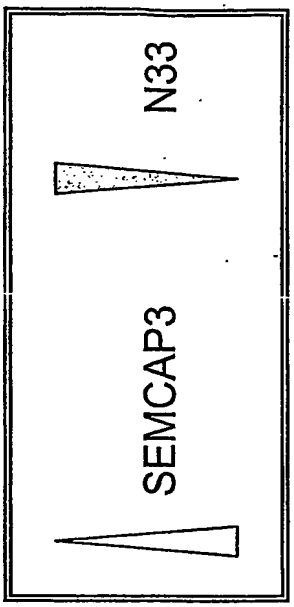
Figure 2 shows that the breakpoint would end *SEMCAP3* transcription after the third exon on the derived chromosome 3 (there would still be one normal chromosome 3 and *SEMCAP3* gene remaining in each nucleus). If transcription occurs on the derived chromosome 3 then the resulting translated protein product would be truncated; lacking part of the first PDZ domain and all subsequent amino acids in the C-terminal direction. It remains to be investigated if the psychiatric disorder in this patient results from *N33* perturbation on one allele, the disruption of *SEMCAP3* on one allele, the generation of an aberrantly functioning truncated SEMCAP3 from one allele or a combination of these.

Pulver *et al.* (1995) detailed schizophrenia linkage to chromosome 3p (albeit telomeric to *SEMCAP3*). However, two further studies have failed to replicate these findings in different populations (Maziade *et al.*, 2001 & Hovatta *et al.*, 1998).

REFERENCES

1. Berrettini W.H. (2000) Genetics of psychiatric disease. Annu. Rev. Med. 51:465 - 79.
2. Lai C.S, Fisher S.E, Hurst J.A, Vargha-Khadem F, Monaco A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 413(6855):519 - 23.
3. Kleinjan D.J, van Heyningen V. (1998) Position effect in human genetic disease. Hum Mol Genet. 7(10):1611 - 8.
4. Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.
5. Wang LH, Kalb RG, Strittmatter SM. (1999) A PDZ protein regulates the distribution of the transmembrane semaphorin, M-SemF. J Biol Chem. 1999 274(20):14137-46.
6. Inagaki S, Furuyama T, Iwahashi Y. (1995) Identification of a member of mouse semaphorin family. FEBS Lett. 370(3):269-72.
7. Pulver A.E, Lasseter V.K, Kasch L, Wolyniec P, Nestadt G, Blouin J.L, Kimberland M, Babb R, Vourlis S, Chen H, et al. (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet. 60(3):252 - 60.
8. Maziade M, Roy MA, Rouillard E, Bissonnette L, Fournier JP, Roy A, Garneau Y, Montgrain N, Potvin A, Cliche D, Dion C, Wallot H, Fournier A, Nicole L, Lavallee JC, Merette C. (2001) A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes. Mol Psychiatry. 6(6):684-93.
9. Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajarvi R, Kokko-Sahin ML, Ekelund J, Lonnqvist J, Peltonen L. (1998) Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a

population-based sampled Finnish family set. Mol Psychiatry. 3(5):452-7.



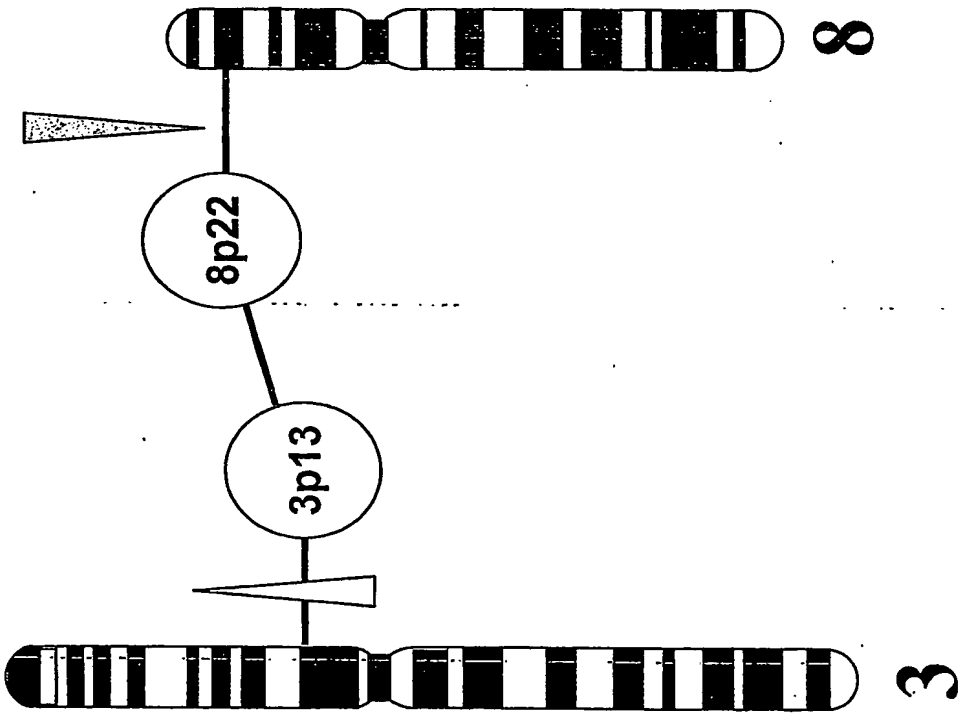


Figure 1

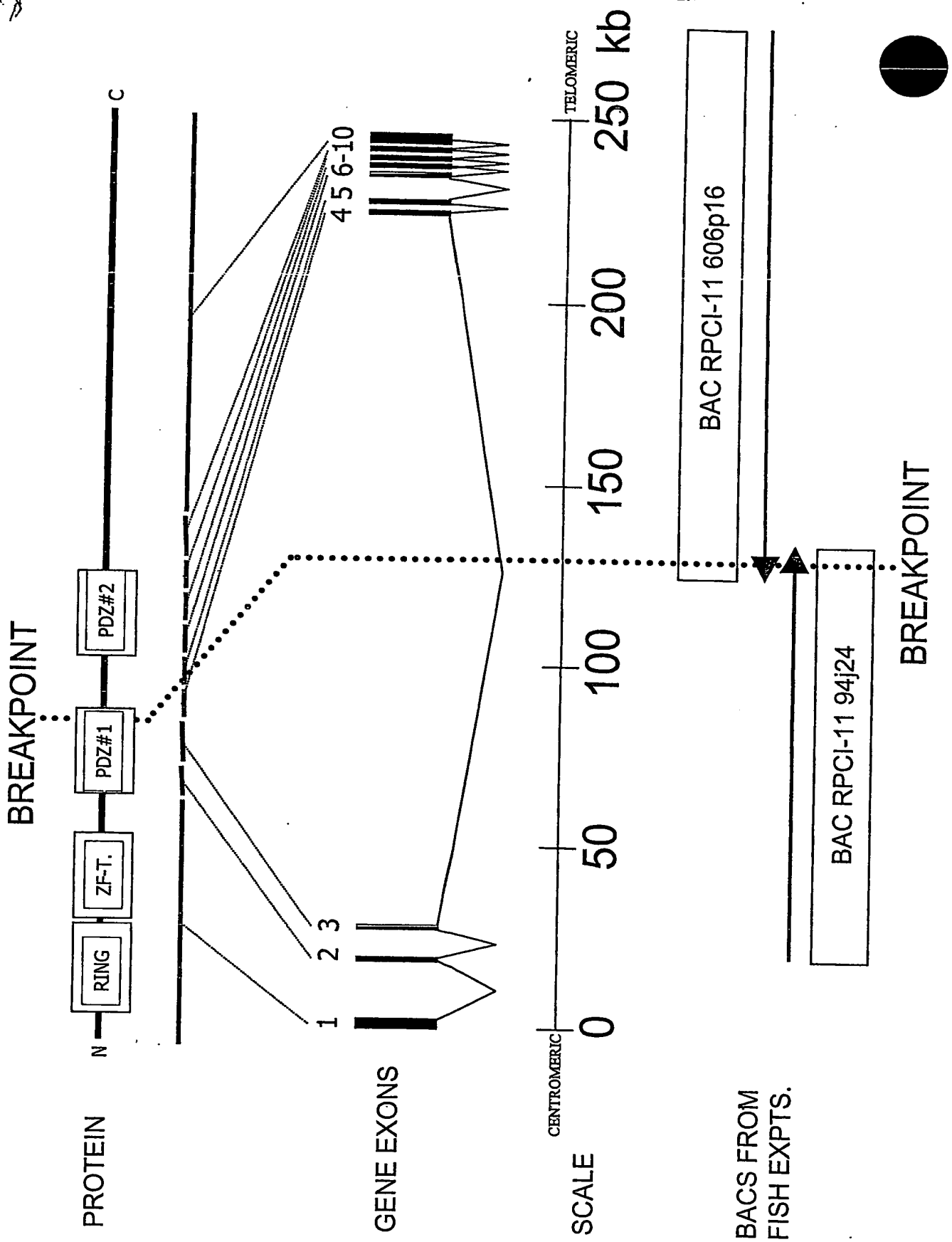

BREAKPOINT
PDZ#2
PDZ#1
ZF-T.
RING
C
N
PROTEIN
1
2 3
4 5 6-10
GENE EXONS
TELOMERIC
CENTROMERIC
0
50
100
150
200
250 kb
SCALE
BAC RPCI-11 606p16
BAC RPCI-11 94j24
BACS FROM FISH EXPTS.
BREAKPOINT

Figure 2.

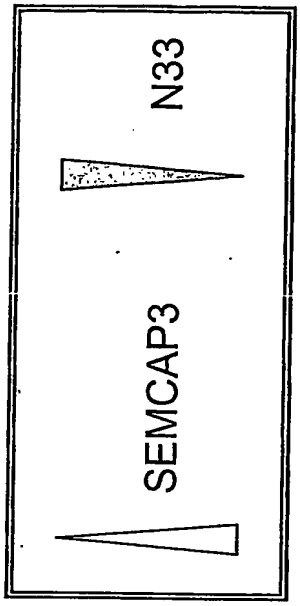
Figure 3

```
   1 AAAACTTCCC CGGGTAGATT CACCCACCGG TCCTGGAAAC CTGCTAAATC CTGAAGGTTC
  61 ACAGAACCTC TGGTCAGAAC TGAAGTTGCA GCCGGAGCTT CCCGCAGGCT CTGTAACTTT
 121 CCCTGGAATG AAATAAATAA ATAAAGACCG TAAGTGCTGA GATAGCGGGC CCCAAGATAT
 181 TTTTAGTCCT CTGCAATCAG CCACTAGAGG AAGGGGGAGG GAGAAGGGAG TAAAAAAGTT
 241 TTGATCCGTT CGGGAAGGGG CTCGAAGAGA ACCCTTGGGA GAAAGCAGTA GCCTCAGCTC
 301 CAAACTCAGC GAGCTTTTCT CGGCTGGCGT TTTGTCTCCT ATAGCGTAGA CTGTAAGAGA
 361 ACAGAAAGGA GTTTCCCGAG AAGATTCAGG CTGGCGTCCT GGGCTGGCCC GTCCCTTCTG
 421 GCGAGCCTCA GTGTCCTCCC ACGCGCTTCT GCCTTCCAGC CTCCTCCCTT TTTCGGGGGG
 481 CTGGCGGGAG GCATCCAAGG CACGATGTAT GTGCGCTCGC GCTCGCGCAA ATACGGCCGG
 541 AGGAGTCCTG TTCCTCGGGC ATTTTCCGAG GAAGTCTGGA TCAATTAGGC TCAGTCCGGG
 601 GAGAGCCAGC GAGCGCGCGG GCGGCGTAGC CGGCCTGTCT GGGCCGCCTC GTGGGGAGGG
 661 AGGGGGCGCC CGGCCGCCCG GCGGCGACCC CGGGGCCTGG CCGCCACCAT GGGCTTCGAG
 721 CTGGACCGCT TCGACGGCGA CGTGGACCCG GACCTGAAGT GCGCGCTGTG CCACAAGGTC
 781 CTGGAGGACC CGCTGACCAC GCCGTGCGGC CACGTCTTCT GCGCCGGCTG CGTGCTGCCC
 841 TGGGTGGTGC AGGAGGGCAG CTGCCCGGCG CGCTGCCGCG GTCGCCTGTC GGCCAAAGAG
 901 CTCAACCACG TCCTGCCGCT CAAGCGCCTT ATCCTCAAGC TGGACATCAA GTGCGCGTAC
 961 GCGACGCGCG GCTGCGGCCG GGTGGTCAAG CTGCAGCAGC TGCCGGAGCA CCTCGAGCGC
1021 TGCGACTTCG CGCCCGCGCG CTGTCGCCAC GCGGGTTGCG GCCAGGTGCT GCTGCGGCGC
1081 GACGTGGAGG CGCACATGCG CGACGCGTGC GACGCGCGGC CAGTGGGCCG CTGCCAGGAG
1141 GGCTGCGGGC TACCCTTGAC GCACGGCGAG CAGCGCGCGG GCGGCCACTG CTGCGCGCGA
1201 GCGCTGCGGG CGCACAACGG CGCGCTCCAG GCCCGCCTGG GCGCGCTGCA CAAGGCGCTC
1261 AAGAAGGAGG CGCTGCGCGC TGGGAAGCGC GAGAAGTCGC TGGTGGCCCA GCTGGCCGCG
1321 GCGCAGCTTG AGCTGCAGAT GACCGCGCTG CGCTACCAGA AGAAATTCAC CGAATACAGC
1381 GCGCGCCTCG ACTCGCTCAG CCGCTGCGTG GCCGCGCCGC CCGGCGGCAA GGGCGAAGAA
1441 ACCAAAAGTC TGACTCTTGT CCTGCATCGG GACTCCGGCT CCCTGGGATT CAATATTATT
1501 GGTGGCCGGC CGAGTGTGGA TAACCACGAT GGATCATCCA GTGAAGGAAT CTTTGTATCC
1561 AAGATAGTTG ACAGTGGGCC TGCAGCCAAG GAAGGAGGCC TGCAAATTCA TGACAGGATT
1621 ATTGAGGTCA ACGGCAGAGA CTTATCCAGA GCAACTCATG ACCAGGCTGT GGAAGCTTTC
1681 AAGACAGCCA AGGAGCCCAT AGTGGTGCAG GTGTTGAGAA GAACACCAAG GACCAAAATG
1741 TTCACGCCTC CATCAGAGTC TCAGCTGGTG GACACGGGAA CCCAAACCGA CATCACCTTT
1801 GAACATATCA TGGCCCTCAC TAAGATGTCC TCTCCCAGCC CACCCGTGCT GGATCCCTAT
1861 CTCTTGCCAG AGGAGCATCC CTCAGCCCAT GAATACTACG ATCCAAATGA CTACATTGGA
1921 GACATCCATC AGGAGATGGA CAGGGAGGAG CTGGAGCTGG AGGAAGTGGA CCTCTACAGA
1981 ATGAACAGCC AGGACAAGCT GGGCCTCACT GTGTGCTACC GGACGGACGA TGAAGACGAC
2041 ATTGGGATTT ATATCAGTGA GATTGACCCT AACAGCATTG CAGCCAAGGA TGGGCGCATC
2101 CGAGAAGGAG ACCGCATTAT CCAGATTAAT GGGATAGAGG TGCAGAACCG TGAAGAGGCT
2161 GTGGCTCTTC TAACCAGTGA AGAAAATAAA AACTTTTCAT TGCTGATTGC AAGGCCTGAA
2221 CTCCAGCTGG ATGAGGGCTG GATGGATGAT GACAGGAACG ACTTTCTGGA TGACCTGCAC
2281 ATGGACATGC TGGAGGAGCA GCACCACCAG GCCATGCAAT TCACAGCTAG CGTGCTGCAG
```

2341 CAGAAGAAGC ACGACGAAGA CGGTGGGACC ACAGATACAG CCACCATCTT GTCCAACCAG
2401 CACGAGAAGG ACAGCGGTGT GGGGCGGACC GACGAGAGCA CCCGTAATGA CGAGAGCTCG
2461 GAGCAAGAGA ACAATGGCGA CGACGCCACC GCATCCTCCA ACCCGCTGGC GGGGCAGAGG
2521 AAGCTCACCT GCAGCCAGGA CACCTTGGGC AGCGGCGACC TGCCCTTCAG CAACGAGTCT
2581 TTCATTTCGG CCGACTGCAC GGACGCCGAC TACCTGGGGA TCCCGGTGGA CGAGTGCGAG
2641 CGCTTCCGCG AGCTCCTGGA GCTCAAGTGC CAGGTGAAGA GCGCCACCCC TTACGGCCTG
2701 TACTACCCTA GCGGCCCCCT GGACGCCGGC AAGAGTGACC CTGAGAGCGT GGACAAGGAG
2761 CTGGAGCTGC TGAACGAAGA GCTGCGCAGC ATCGAGCTGG AGTGCCTGAG CATCGTGCGC
2821 GCCCACAAGA TGCAGCAGCT CAAGGAGCAG TACCGCGAGT CCTGGATGCT GCACAACAGC
2881 GGCTTCCGCA ACTACAACAC CAGCATCGAC GTGCGCAGAC ACGAGCTCTC AGATATCACC
2941 GAGCTCCCGG AGAAATCCGA CAAGGACAGC TCGAGCGCCT ACAACACAGG CGAGAGCTGC
3001 CGCAGCACCC CGCTCACCCT GGAGATCTCC CCCGACAACT CCTTGAGGAG AGCGGCGGAG
3061 GGCATCAGCT GCCCGAGCAG CGAAGGGGCT GTGGGGACCA CGGAAGCCTA CGGGCCAGCC
3121 TCCAAGAATC TGCTCTCCAT CACGGAAGAT CCCGAAGTGG GCACCCCTAC CTATAGCCCG
3181 TCCCTGAAGG AGCTGGACCC CAACCAGCCC CTGGAAAGCA AAGAGCGGAG AGCCAGCGAC
3241 GGGAGCCGGA GCCCCACGCC CAGCCAGAAG CTGGGCAGCG CCTACCTGCC CTCCTATCAC
3301 CACTCCCCAT ACAAGCACGC GCACATCCCG GCGCACGCCC AGCACTACCA GAGCTACATG
3361 CAGCTGATCC AGCAGAAGTC GGCCGTGGAG TACGCGCAAA GCCAGATGAG CCTGGTGAGC
3421 ATGTGCAAGG ACCTGAGCTC TCCCACCCCG TCGGAGCCGC GCATGGAGTG GAAGGTGAAG
3481 ATCCGCAGCG ACGGGACGCG CTACATCACC AAGAGGCCCG TGCGGGACCG CCTGCTGCGG
3541 GAGCGCGCCC TGAAGATCCG GGAAGAGCGC AGCGGCATGA CCACCGACGA CGACGCGGTG
3601 AGCGAGATGA AGATGGGGCG CTACTGGAGC AAGGAGGAGA GGAAGCAGCA CCTGGTGAAG
3661 GCCAAGGAGC AGCGGCGGCG GCGCGAGTTC ATGATGCAGA GCAGGTTGGA TTGTCTCAAG
3721 GAGCAGCAAG CAGCCGATGA CAGGAAGGAG ATGAACATTC TCGAACTGAG CCACAAAAAG
3781 ATGATGAAGA AGAGGAATAA GAAAATCTTC GATAACTGGA TGACGATCCA AGAACTCTTA
3841 ACCCACGGCA CAAAATCCCC GGACGGCACT AGAGTATACA ATTCCTTCCT ATCGGTGACT
3901 ACTGTATAAT TTTCACTTCT GCATTATGTA CATAAAGGAG ACCACTACCA CTGGGGTAGA
3961 AATTCCTGCC TCGTTCAATG CGGCAAGTTT TTGTATATAA GATAAGTACG GTCTTCATGT
4021 TTATAGTCCA AATTTGCAAA CCCTACAACT CTGGGTGTCA TAGGTCTATT TTAAGGGAAG
4081 AGAGAGAAAA ACACCCTTAC TATCTTGGAA GGCAATATTA ACAAACAGAG CTTTTTTCAA
4141 ATAGCAATTG TACTTTTCTA CCTGTACCCT TTTACATAAA GTGTTTAAAT TTCAGAAAGA
4201 TCTTTTATTA AGCATACTTT CACAGAATAA CTTGTTTAAA CTATATTCAT ATAAAAAAGT
4261 TAAACACGCT TTTTTTCCTG CCTAAAACAC AAATACAACT GCCAGTATGT ATTTTTAATG
4321 GAACCCTATT TTATAATGGT ACGTTACTGA ATGTGTTTCA TATGCGTGAC CGTTAAGATA
4381 TTATCATTTA GGTGAAGGTT TCAACTCAAA ACCACCCAAC CCGGTGGTTA ACGATTTAAT
4441 ACATATAACC AAACCGGCAG CGTTTAGAGT TGGGATATAC ATTTAAACAT TTTCCTGGTT
4501 AAGGTTCCCA AGAGAGTGTA AAGGTTTTAG CAGAAAGCAA AATATCTTGC ATCTTTATGG
4561 AAGTTTAAAG CATGTTTGCA AATATTGCAG CCCATTGAAA GAATTTGCAT GTACAGGAAA
4621 GTTGTGGATG GAGACGGTTT GTGGAATTTT AAGTGCTCAT TGTAGTAAAC TTTTGCTTTG
4681 TAGATTTGAA GGTACAGACT TATACAGGCA AGTTCACAAA ATCATGATTA GTTACAAACA
4741 GTAAAATGAA GTTAAAATAA ATTATTATTT TCT

Figure 4

```
   1 MGFELDRFDG DVDPDLKCAL CHKVLEDPLT TPCGHVFCAG CVLPWVVQEG SCPARCRGRL
  61 SAKELNHVLP LKRLILKLDI KCAYATRGCG RVVKLQQLPE HLERCDFAPA RCRHAGCGQV
 121 LLRRDVEAHM RDACDARPVG RCQEGCGLPL THGEQRAGGH CCARALRAHN GALQARLGAL
 181 HKALKKEALR AGKREKSLVA QLAAAQLELQ MTALRYQKKF TEYSARLDSL SRCVAAPPGG
 241 KGEETKSLTL VLHRDSGSLG FNIIGGRPSV DNHDGSSSEG IFVSKIVDSG PAAKEGGLQI
 301 HDRIIEVNGR DLSRATHDQA VEAFKTAKEP IVVQVLRRTP RTKMFTPPSE SQLVDTGTQT
 361 DITFEHIMAL TKMSSPSPPV LDPYLLPEEH PSAHEYYDPN DYIGDIHQEM DREELELEEV
 421 DLYRMNSQDK LGLTVCYRTD DEDDIGIYIS EIDPNSIAAK DGRIREGDRI IQINGIEVQN
 481 REEAVALLTS EENKNFSLLI ARPELQLDEG WMDDDRNDFL DDLHMDMLEE QHHQAMQFTA
 541 SVLQQKKHDE DGGTTDTATI LSNQHEKDSG VGRTDESTRN DESSEQENNG DDATASSNPL
 601 AGQRKLTCSQ DTLGSGDLPF SNESFISADC TDADYLGIPV DECERFRELL ELKCQVKSAT
 661 PYGLYYPSGP LDAGKSDPES VDKELELLNE ELRSIELECL SIVRAHKMQQ LKEQYRESWM
 721 LHNSGFRNYN TSIDVRRHEL SDITELPEKS DKDSSSAYNT GESCRSTPLT LEISPDNSLR
 781 RAAEGISCPS SEGAVGTTEA YGPASKNLLS ITEDPEVGTP TYSPSLKELD PNQPLESKER
 841 RASDGSRSPT PSQKLGSAYL PSYHHSPYKH AHIPAHAQHY QSYMQLIQQK SAVEYAQSQM
 901 SLVSMCKDLS SPTPSEPRME WKVKIRSDGT RYITKRPVRD RLLRERALKI REERSGMTTD
 961 DDAVSEMKMG RYWSKEERKQ HLVKAKEQRR RREFMMQSRL DCLKEQQAAD DRKEMNILEL
1021 SHKKMMKKRN KKIFDNWMTI QELLTHGTKS PDGTRVYNSF LSVTTV
```



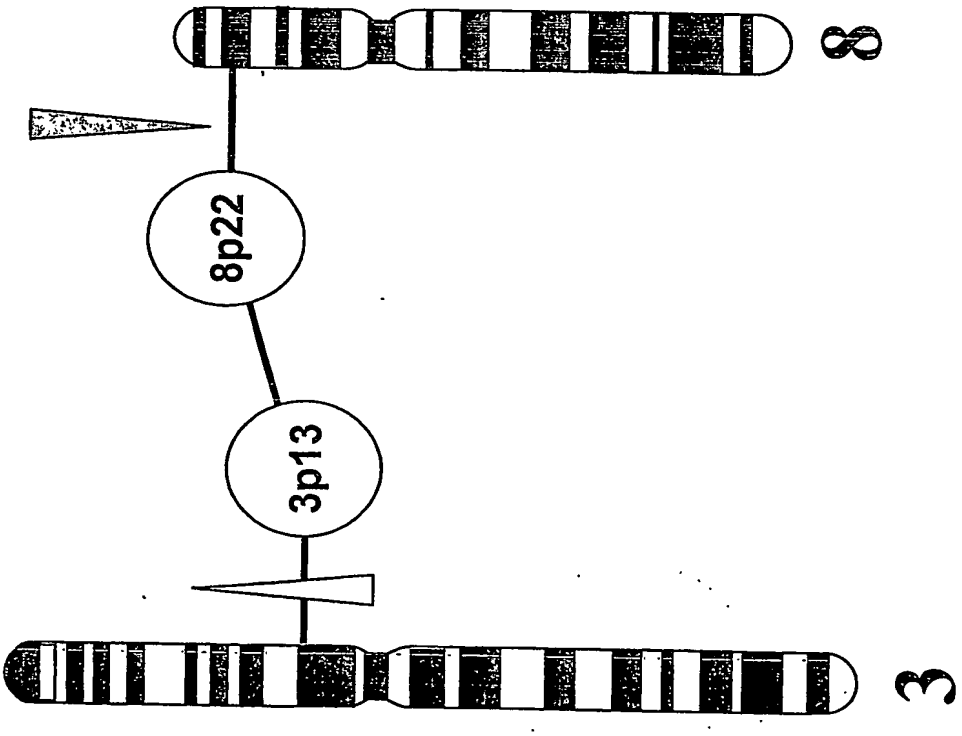


Figure 1